being classified as mentally retarded. Since the tests had been normed at different dates, they produced radically higher or lower IQs until the scores were adjusted to take into account which tests had been normed recently and which had been normed as much as 30 years before. Since 1998, when the American Psychological Association issued *The Rising Curve: Long-Term Gains in IQ and Related Measures* (Neisser, 1998), no scholar published in a first-line journal has ignored the relevant data on IQ gains over time. Recent examples include *IQ and the Wealth of Nations* (Lynn & Vanhanen, 2002) and the study headed by the noted scholar Naomi Breslau, tracking the relationship between low birth weight, social disadvantage, and IQ during the years of school attendance (Breslau, Dickens, Flynn, Peterson, & Lucia, in press).

The problem of adjusting IQ scores across nations and across different kinds of tests is complex. The rate of gain varies by country—for example, from America to Israel to Norway—and by test—for example, from matrix tests to purely verbal IQ tests (Flynn, 2006). Those who deal with the scores of Americans on Wechsler and Stanford–Binet (SB) IQ tests are fortunate in that the rate of gain is roughly uniform. The standard practice is to deduct 0.3 IQ points per year (3 points per decade) to cover the period between the year the test was normed and the year in which the subject took the test. The evidence for that value is discussed in a subsequent section of this article. For now, take it for granted that the relevant rate of gain is 0.3 IQ points per year.

Why do scholars make these adjustments? Imagine you are running an educational project whose objective is to raise the IQs of young children in a depressed area. You give one IQ test before they enter the program at age 4 and another when they leave it at age 6. They appear to have gained 10 IQ points. Then you notice that the first test was normed 5 years before you administered it, the second 25 years before. Thanks to IQ gains over time, the entry test was normed on a better performing sample. This is clear because studies in which the same children took *both tests* reveal the following pattern: They scored 100 on the recent test, which shows that their performance was merely average; and yet, the very same children receive an IQ of 106 when scored on the norms of the older test. In other words, a sample of children drawn 20 years ago set a lower standard than children do today. The 6 points higher IQ is purely a matter of being scored against an earlier and lower performing group. Naturally, what concerns you is how much the experimental children really improved in their IQ performance. Therefore, you deduct 6 points from the IQ scores they received when they left the program: 0.3 points for each of the 20 years between the norming of the two tests. This shows that the experimental children really gained only 4 points from participating in your project, not the 10 points the unadjusted IQ scores suggested.

Failure to adjust the scores is to take flight from reality. Suppose you are coaching an athlete who aspires to qualify for the Olympic high jump. He jumps 6 ft 6 in., and you assure him that he will qualify. He replies: "But that was the standard in 1985. Since then, performances have improved, and today, I have to jump 7 feet to qualify. You are judging my performance in terms of the norms of yesterday rather than today." He would do well to hire a new coach.

Clinical and school psychologists have been slow to recognize the problems posed when a new IQ test replaces an older one (for full names of all tests, see Table 1). Kanaya, Scullin, and Ceci (2003) found that after the Wechsler Intel-

ligence Scale for Children—Third Edition (WISC–III) was published in 1991, there was an immediate rise in the number of children being classified as mentally retarded. This was because some school psychologists had begun to use the new test (normed in 1989) rather than the older WISC–R (normed in 1972). Thanks to IQ gains over time, children in the retardate range who took the new test were averaging 5.6 IQ points lower than those who were still taking the old test. Note, by the way, how close 5.6 points is to the 5.1 points entailed by a gain of 0.3 points per year over the 17 years between the selection of the two standardization samples. Failure to adjust for the difference between the two tests meant that beginning in 1992, whether a child was classified as mentally retarded was likely to depend on whether school psychologists had used up their stock of the old WISC–R test or had purchased the new WISC–III test.

The AAMR (2002) has cited the Flynn effect as evidence that it "is critically important to use standardized tests with the most updated norms" (p. 56) and notes score changes "between successive revisions" (p. 59) of IQ tests. It does not discuss the mechanics of translating scores on an old version of the WISC into scores on a current one. After all, if you are a practicing school psychologist and see a child with an obsolete IQ score, the obvious remedy is to administer the current test. However, when a court wants to establish whether defendants in their 40s showed signs of MR before the age of 18, they cannot regress them to childhood and administer a new test. If the tests taken had current norms at the time they were taken, all is well. But if the norms were obsolete, all that can be done is to reassess the IQ scores recorded in the light of IQ gains over time.

The California court cited in *People v. Superior Court* (2005) in California goes further than I would in asserting that the Flynn effect seems to be generally accepted in the clinical field. However, the comments of Trowbridge (2003) and Frumkin (2003) on how to deal with expert testimony about IQ scores show that the formula of deducting 0.3 points per year is making headway in capital cases.

IQ Gains and Capital Cases

The case for taking IQ gains into account in capital cases is best presented by noting the consequences of failing to do so. Consider a boy whose IQ was tested twice during his school years. In 1973, at age 6, he scored 75 on the WISC, which was normed in 1947–1948 and whose norms were 25.5 years out of date. In 1975, at age 8, he scored 68 on the new WISC–R, which was normed in 1972 and whose norms were only 3 years out of date. Both scores show that this boy was mentally retarded: The score at age 6 is deceptive because it was inflated by norms 25.5 years old, which is to say by over 7 IQ points. When those points are deducted, both scores become 68. The boy's true IQ did not alter between ages 6 and 8: All that changed were the norms.

Now imagine we split this example into two boys who are today subject to the death penalty. At school, they performed identically on IQ tests. But one is executed because he was unlucky enough to take the WISC and scored 75 when measured against its obsolete norms; the other is excused as mentally retarded because he was lucky enough to take the WISC–R and scored 68 when measured against its relatively current norms. Failure to adjust IQ scores in the light of IQ

gains over time turns eligibility for execution into a lottery—a matter of luck about what test a school psychologist happened to administer.

IQ Scores and Criteria of MR

Prosecution experts sometimes claim that psychologists are still evaluating the significance of the Flynn effect. This is true. But it does not mean that we should wait for some distant day before applying our knowledge of the rate of IQ gains to adjust the IQ scores of defendants in current cases.

Sooner or later, we must decide whether retardate IQ gains represent enhanced ability to deal with everyday life and enhanced ability to make decisions for which one is culpable. If IQ gains mean that fewer and fewer Americans lack those capabilities, the AAMR criterion of MR should be altered. Perhaps it should be set at three standard deviations below the population mean (an IQ of 55 against current norms) rather than two standard deviations (an IQ of 70). I will argue in favor of the status quo.

First, we now have data that compare the performance of samples used to standardize the Vineland Adaptive Behavior Scales at two different times (Sparrow, 2006). For ages 7 to 18, subjects who took both tests actually found the 1984 norms more difficult to meet than current norms. That is, they received an overall Adaptive Behavior Composite of only 95.0 on the old test and one of 98.4 on the new test ($SD = 15$). This seems to indicate that American schoolchildren actually lost ground in terms of adaptive behavior since 1984.

However, their scores on the Communication and Socialization subtests were similar on the two versions. The lost ground was almost entirely on the Daily Living Skills subtest. The 1984 version of that subtest contains obsolete skills that would deflate the scores of contemporary children (items such as "sews or hems clothes," "makes own bed," and "uses a pay telephone"). The most judicious conclusion is that American children have marked time. In sum, they made no gains in terms of adaptive behavior during a period in which they made large IQ gains. This result suggests that IQ gains over time do not mean that fewer and fewer children find it difficult to cope with everyday life.

Second, IQ gains over time show a peculiar pattern (Flynn, 2006): They are huge on some WISC subtests, such as Verbal Similarities and Block Design, and virtually nil on others subtests, such as Vocabulary and General Information. I believe that IQ gains are measuring real skill gains but that these gains do not constitute global intelligence gains. The rise in Verbal Similarities scores signals the penetration of the scientific ethos into American minds. If asked what dogs and rabbits have in common, the prescientific children of 1900 tended to say, "you use dogs to hunt rabbits," a satisfactory concrete operational answer but one for which they would score zero. The postscientific children of today will say that they are both mammals, a formal taxonic answer for which they will score full marks. The visual culture of today has upgraded performance on Block Design. And it is important that America has more citizens who can see the world through scientific spectacles and can deal with the Internet. But once again, these gains do not mean that people of low intelligence are any better at concrete operational decisions in everyday life. They certainly have no greater fund of general

information to inform those decisions and no greater core vocabulary to verbalize them.

Third, assume I am mistaken and that IQ gains in fact signal some sort of global intelligence gain. The main environment to which a low-IQ person must adapt is the expectations and behavior of other people. The more intelligent the average child becomes, the greater the cognitive demands we make on all schoolchildren. The more intelligent the average worker becomes, the less likely an employer will be to hire those who need a foreman to stand over them and tell them what to do. And what of the culpability of the low-IQ person? Cases often involve a ringleader of normal intelligence who has recruited someone of low IQ to participate in an armed robbery. The ringleader kills someone, but the low-IQ person present is an accessory. If everyone is getting brighter, low-IQ persons may be a bit more resistant to being manipulated, but the average person is also more persuasive in attempting to manipulate them.

For example, I was recently asked to comment on a case in which the companions of the defendant's youth had submitted affidavits. One testified that when the defendant and he were both 16 years of age, the companion pointed out a car and said: "That is Mrs. Smith's car. She and I are friends and she said I could borrow her tires. Would you go over and take them off and bring them to me?"

I believe that jurists should allow the AAMR to decide issues concerning what IQ score should serve as a criterion of MR. Whatever its decision, the salient point is this: No matter whether the criterion is set at an IQ of 55 or 70 or 85, the defendant must be assessed against current norms and not obsolete norms that inflate his or her score. Otherwise, one person will meet the criterion of mental retardation, and another person will be judged not to have done so, purely because one took a test with current norms and the other took a test with obsolete norms. No matter what the criterion, who meets it must not be a matter of chance. Someone who took the WISC and reaped an "unearned" bonus of 7 IQ points must not be executed, whereas someone who was lucky enough to take the WISC–R is exempt. The only way to avoid such inequity is to adjust IQ scores in the light of IQ gains over time.

IQ Gains and Evidence

Table 1 presents evidence showing that IQ gains, at least in the United States on Wechsler–Binet tests, have proceeded at the rate of 0.3 points per year or 3 points per decade. Prior to the SB–5, the SB standard deviation was set at 16 IQ points, which means that 68 (two standard deviations below the mean) was the cutting line for MR. In Table 1, all scores inclusive of older SB scores have been adjusted to a common metric of $SD = 15$, so that IQ gains can be calculated without ambiguity. All test comparisons are derived by analysis of data provided in the Wechsler and SB publishers' manuals. In every instance, they gave the same subjects two tests normed some years apart. The comparisons of WISC and WAIS are based on subjects ages 16–17, the age at which the norms of the two tests overlap. IQ gains from one test to the next have been divided by the number of years that separated the norming of the two tests. This gives us 12 estimates of the rate of IQ gains over time covering the period 1972 to 2002.

When the 12 estimates are averaged, they yield a rate of gain of 0.300 points per year plus or minus less than one hundredth of a point. The slight ambiguity

Table 1
Using Recent IQ Gains to Evaluate the WAIS–III

Tests compared	Gains	Period (yrs)	Rate
Comparisons without WAIS–III			
WISC–R (1972) & WAIS–R (1978)[a]	+0.90	6	+0.150
SB–LM (1972) & SB–4 (1985)[b]	+2.16[c]	13	+0.166
WISC–R (1972) & SB–4 (1985)[d]	+2.95[c]	13	+0.227
WISC–R (1972) & WISC–III (1989)[e]	+5.30	17	+0.312
WAIS–R (1978) & SB–4 (1985)[f]	+3.42[c]	7	+0.489
SB–4 (1885) & SB–5 (2001)[g]	+2.77[c]	16	+0.173
WISC–III (1989) & WISC–III/IV[h] (2001.75)[i]	+4.23	12.75	+0.332
WISC–III (1989) & SB–5 (2001)[j]	+5.00	12	+0.417
Comparisons with WAIS–III			
WAIS–R (1978) & WAIS–III (1995)[k]	+2.90	17	+0.171
WISC–III (1989) & WAIS–III (1995)[l]	−0.70	6	−0.117
WAIS–III (1995) & SB–5 (2001)[m]	+5.50	6	+0.917
WAIS–III (1995) & WISC–IV (2001.75)[n]	+3.10	6.75	+0.459
Comparisons with WAIS–III scores lowered by 2.34 IQ points			
WAIS–R (1978) & WAIS–III (1995)	+5.24	17	+0.308
WISC–III (1989) & WAIS–III (1995)	+1.64	6	+0.273
WAIS–III (1995) & SB–5 (2001)	+3.16	6	+0.527
WAIS–III (1995) & WISC–IV (2001.75)	+0.76	6.75	+0.113

Note. All dates assigned to tests refer to the date at which the test was normed. This is what is relevant, of course, not the date when the test was published. Another date that practitioners might like to have is that for the norming of the WISC: from 1947 to 1948. The data from all sources listed below have been analyzed and adapted to elicit comparisons.
[a]Wechsler Intelligence Scale for Children—Revised (WISC–R) & Wechsler Adult Intelligence Scale—Revised (WAIS–R): Wechsler (1981), Table 18. [b]Stanford–Binet (SB) LM & SB–4: Thorndike et al. (1986), Table 6.6. [c]Prior to the SB–5, the SB standard deviation was set at 16 IQ points, rather than the usual 15 points. The above estimates are all based on scores adjusted to a common metric of $SD = 15$. [d]WISC–R & SB–4: Thorndike et al. (1986), Table 6.7. [e]WISC–R & WISC–III: Flynn (1998), Table 1. [f]WAIS–R & SB–4: Thorndike et al. (1986), Table 6.9. [g]SB–4 & SB–5: Roid (2003), Table 4.1. [h]The alert reader will have noticed the peculiar label given in the comparison of the WISC–III with its successor test, namely, that the latter is called the WISC–III/IV. The WISC–IV dropped 5 of the 10 subtests of the WISC–III, and this rendered the two noncomparable in terms of estimating the rate at which the WISC–III had become obsolete. Fortunately, the Psychological Corporation had collected special data (see Note i below) that offered a solution. Flynn and Weiss (of the Psychological Corporation) used that data to simulate how the WISC–IV standardization sample would have performed on the unaltered WISC–III. They found that IQ scores would have been at least 1.33 points lower than the WISC–IV yielded—thus, the odd label WISC-III/IV, which refers to using a WISC–IV sample to assess norms for a test like the WISC–III. [i]WISC–III & WISC–III/IV: Flynn & Weiss (2006). The estimate given is the midpoint of the range of estimates for this pair of tests. [j]WISC–III & SB–5: Roid (2003), Table 4.6. [k]WAIS–R & WAIS–III: Wechsler (1997), Table 4. [l]WISC–III & WAIS–III: Wechsler (1997), Table 4.3. [m]WAIS–III & SB–5: Roid (2003), Table 4.7. [n]WAIS–III & WISC–IV: Psychological Corporation (2003), Table 5.12.

has to do with whether WAIS–III scores should be allowed to stand or whether they must be adjusted because its norms are suspect. When those who have committed capital crimes are tested postconviction, and they always are, a recognized test with norms for adults must be used. The only ones available are the SB–5 and the WAIS–III, with the latter the most popular choice. Therefore, any problem with its norms will have serious consequences.

The WAIS–III problem arises out of its effects within the total matrix of comparisons. Assume a sequence in which candidates average an IQ of 106 on Test A (normed 1980), 103 on Test B (normed 1990), and 100 on Test C (normed 2000). That would indicate a steady IQ gain of 0.3 points per year, as the norms get tougher to exceed over time. However, imagine that Test B disrupts this pattern wherever it appears—for example, candidates average 106 on Test A, 106 on Test B, and 100 on Test C. The effect is that wherever Test B occurs as the second test in a combination, it deflates the rate of gain to zero: The combination of A followed by B gives no gains over 10 years. And wherever Test B occurs as the first test in a combination, it inflates the rate of gain to improbable levels: The combination of B followed by C gives a huge gain of 6 points over 10 years. We would have a choice between concluding that the rate of gain was wildly eccentric and concluding that there was something wrong with the norms of Test B, namely, that its standardization sample was below average and thus inflated IQs by 3 points.

In accord with this logic, Table 1 is divided into three sections. The first section isolates the eight comparisons in which the WAIS–III does not appear. These give an average rate of gain of 0.283 points per year with a relatively narrow range of estimates (from 0.150 points per year to 0.489 points), particularly if one takes into account that no one estimate should be trusted. The second section isolates the four comparisons in which the WAIS–III does appear. These give an average rate of gain of 0.358 points per year with a huge range of estimates, running all the way from a "minus gain" (a loss of 0.117 points per year) to a gain of almost an IQ point per year—the latter far too large to be taken seriously. Looking at what is happening in those four comparisons, it is clear that the WAIS–III inflated IQ scores by about 2.34 points even at the time it was normed. That is, its normative sample was a bit substandard compared with other IQ tests, and therefore, its average performance was easier to beat. A performance that would be an IQ of 100 on the others becomes an IQ of 102.34 on the WAIS–III.

The third section of the table delivers the proof. When all WAIS–III scores are lowered by 2.34 points, the four comparisons in which the WAIS–III appears fall into line. They average rate of gain of falls from 0.358 points per year to 0.305 points per year, which is virtually identical to the rate yielded by the average of all 12 comparisons. And the wide range of estimates narrows to virtual parity with comparisons in which the WAIS–III does not appear.

In sum, when its scores are left unadjusted, the WAIS–III is behaving just like Test B. Whenever it is the later test, it produces tiny rates of gain or even losses (note the values of 0.171 and −0.117 in the second section of Table 1)—by inflating IQs. Whenever it is the earlier test, it produces huge rates of gain (note the values of 0.917 and 0.459 in the same section)—also by inflating IQs. When the WAIS–III was published in 1997, the results from that time to the present were

of course unavailable. Therefore, it prompted speculation that the rate of IQ gains might have diminished (Flynn, 1998); that speculation is now seen to be amiss.

Table 1 shows only that the WAIS–III's norms are atypical. However, the alternative to considering that its norms are suspect is to doubt the norms of all four of the other tests that occur in comparison with it: the WAIS–R, the WISC–III, the SB–5, and the WISC–IV. It is logically possible that rather than the WAIS–III team *selecting* a substandard sample, the architects of all these other tests selected elite samples. If that is so, all tests but the WAIS–III are deflating IQs and many are being labeled as mentally retarded who should not qualify. But the odds are 15 to 1 against it. Table 1 suggests a general rule: Deduct 0.3 IQ points per year from the scores of defendants for every year that passed between when the test was normed and when the test was taken. It also suggests an addendum: Deduct an extra 2.34 points from WAIS–III scores on the grounds that it gave inflated IQs even in the year in which it was normed.

IQ Gains and Low-IQ Subjects

Before the general rule is applied, the question naturally arises as to whether IQ gains at low-IQ levels are similar in magnitude to those at the mean. If not, the rule of 0.3 points of obsolescence per year might have to be modified for those on the borderline of retardation—for example, those who score 75 on an obsolete test and need a deduction of 5 points to qualify as mentally retarded. Flynn (1985, 1998) and Spitz (1989) have collected large bodies of data at low-IQ levels. These have been supplemented by studies done by the testers themselves and by Kanaya et al. (2003).

Table 2 is not designed to allow conversion of IQs from one test to another but to use combinations of tests and their standardization samples to estimate the rate at which the norms of preceding tests became obsolete. The most important column is that headed "Predicted versus actual" which compares the predicted obsolescence (given by the general rule) with the actual value yielded by the data. The predicted obsolescence is based on taking the years between the standardization *of each pair* of tests and multiplying by 0.3 IQ points. Using the WISC tests as an example yields the following results: The WISC and WISC–R were normed 24.5 years apart, so multiplying by 0.3 gives 7.35 points as the predicted obsolescence; the WISC–R and WISC–III were normed 17 years apart, which gives 5.10 points; the WISC–III and WISC–IV were normed 12.75 years apart, which gives 3.825 points; and so forth for all of the other test included in the table.

Table 2 shows that the WISC tests behave with remarkable consistency. The data sets are large, and at every IQ level from 55 to 100, the predicted and actual values are very close. If anything, at low-IQ levels, the original WISC norms have become obsolete a bit faster (at about 0.4 points per year) than the rate set by the general rule (0.3 points).

But why are the WAIS and WAIS–R data different? Particularly at very low-IQ levels, we find what appear to be IQ losses rather than IQ gains. Throughout the world, in the relevant period (1953–1954 to 1978), IQ gains have occurred on all manner of tests, and the only controversy has been whether the obvious gains in the bottom half of the curve were matched by gains in the top half

Table 2
IQ Differences by IQ Level: Predicted Versus Actual Score Differences

WISC & WISC–R[a]		WISC–R & WISC–III		WISC–III & WISC–III/IV[b]	
IQ differences by level	Predicted vs. actual	IQ differences by level	Predicted vs. actual	IQ differences by level	Predicted vs. actual
90–100 = 81–91	7.35/8.52	90–110 = 86–104	5.10/5.02	100 = 96–98	3.825/4.33
80–90 = 71–81	7.35/8.89	80–90 = 75–85	5.10/5.33	85 = 81–84	3.825/3.83
70–80 = 61–71	7.35/8.52	75–80 = 70–75	5.10/4.78	70 = 65–70	3.825/3.83
55–70 = 45–60	7.35/9.70	60–75 = 53–68	5.10/6.77	55 = 49–56	3.825/3.83

WAIS & WAIS–R		WAIS-R & WAIS–III	
IQ differences by level	Predicted vs. actual	IQ differences by level	Predicted vs. actual[c]
80–100 = 73–93	7.35/6.90	100 = 96–98	5.10/5.34 (3.00 + 2.34)
[70–80 = 67–77][d]	nil/3.47	85 = 81–83	5.10/5.34 (3.00 + 2.34)
65–70 = —	—	70 = 65–69	5.10/5.34 (3.00 + 2.34)
[55–65 = 56–66][d]	nil/−1.23	55 = 49–54	5.10/5.84 (3.50 + 2.34)
[40–55 = 46–61][d]	nil/−6.28	—	—

Note. Dashes indicate that no data were available. Numbers for each comparison (running from higher to lower IQs) are as follows: WISC & WISC–R, 326, 415, 259, 170; WISC–R & WISC–III, 652, 1055, 215, 108; WISC–III & WISC–III/IV, 244 (only total for all subjects at all IQ levels is given); WAIS & WAIS–R, 2 studies, 1 study, —, 3 studies, 1 study. The average study had 52 subjects and none had under 20; WAIS–R & WAIS–III, 192 (only total for all subjects at all IQ levels is given). The data from the following sources have been analyzed and adapted: WISC & WISC–R: Flynn, 1985, Table 2; WISC–R & WISC–III: Zimmerman & Woo-Sam, 1997, Table 1, and Kanaya et al., 2003; WISC–III & WISC–III/IV: Psychological Corporation, 2003, Table 5.9; WAIS & WAIS–R: Spitz, 1989, Table 1; WAIS–R & WAIS–III: Wechsler, 1997, Table 4.1. WISC = Wechsler Intelligence Scale for Children; WAIS = Wechsler Adult Intelligence Scale.
[a]WISC–R scores adjusted to simulate scoring against the White members of the normative sample, thus allowing comparison with the WISC whose sample was White. [b]See Table 1 Note h: Flynn and Weiss found that (at least) 1.33 points had to be deducted from WISC–IV IQs to equate them with WISC–III IQs, so that was added to the WISC–III & WISC–IV differences at all levels—to produce the "WISC–III & WISC–III/IV" differences. [c] WAIS–III actual scores have been treated as being 2.34 points too high, so that amount was added to the "actual" differences between it and the earlier test. [d]The bracketed scores yield no real comparisons due to differences in scoring conventions.

(Flynn, 2006). Can low-IQ Americans above the age of 16 (the ages covered by the WAIS) be the only group in the world that differs?

Table 3 provides the answer by comparing the WISC–R and WAIS–R. Because these tests were normed only 6 years apart, one would expect WISC–R scores (1972) to be only slightly higher than WAIS–R scores (1978)—and they are higher at the mean. But low-IQ levels are wildly out of step. At a WISC IQ of 46, we get a reverse IQ gain (a loss) of fully 15.65 points, which gradually erodes as we go up through IQs of 64 and 73 and 81 and finally turns into a gain at the mean. This peculiar pattern suggests that the score differences at low-IQ levels measure not norms becoming obsolescent thanks to IQ gains but disparate scoring conventions.

Let us look at the test manuals. The WISC–R makes it difficult to get even a low IQ (Wechsler, 1974, pp. 150, 152). An IQ of 48 requires a standard score of 26, which entails a raw score of 147, or at least 74 items correct (some items give 2 points for a correct answer, but others give only 1). The WAIS–R is notorious for what came to be called the "tree-stump" phenomenon. An IQ of 48 requires a standard score of 13, but that can be "earned" with zero raw score points (Wechsler, 1981, pp. 93, 142), which is to say that an inert subject who answered not one item correctly qualifies. The WISC–R items in question may appear easy, but they are not easy for those three standard deviations below average (the bottom one eighth of 1% of the population), and getting them right is not nearly as easy as getting no items right at all. In Table 3, the lowest IQ group certainly did better on the WAIS–R than the 45.90 they scored on the WISC–R. But they had to score higher: No items correct by anyone would have given them 48. That they scored 61.55 simply illustrates the potency of the tree-stump phenomenon.

There is only one hypothesis that could justify using WAIS–R data at low-IQ levels to calculate obsolescence. The WISC to WISC–R data combine scores for all ages from 5 to 15 years. Treating them collectively, a rate of gain of 0.4 points per year (see Table 2) would give them a 2.4-point gain over 6 years. But imagine

Table 3

Evidence That WAIS–R Scores Imply a Huge Variation of IQ Gains at Low-IQ Levels (the "Tree-Stump" Phenomenon)

WISC–R[a]	WAIS–R[a]	Difference	% difference intact
45.90	61.55	−15.65	100
63.67	71.05	−7.38	47
72.73	77.23	−4.60	29
81.13	82.70	−1.57	10
(91)	(91)	nil	nil
101.90	101.00	+0.90	−6

Note. The data from all sources have been analyzed and adapted to elicit comparisons. Sources: Spitz, 1989, Table 2, and Wechsler, 1981, p. 48. WAIS–R = Wechsler Adult Intelligence Scale—Revised; WISC–R = Wechsler Intelligence Scale for Children—Revised.

[a]WISC–R & WAIS–R comparison studies (numbers from low to high IQs): 2 studies, 2 studies, 3 studies, 1 study, —, 1 study. The study of those with average IQs (WISC–R = 101.90) had 80 subjects. The remaining 8 had a total of 242 for an average of 30, and none had fewer than 20.

there were large age differences. If gains by 5-year-olds turned into losses by age 15, then the WISC values and the WAIS–R values for ages 16–17 could be *reconciled*. To test this, we must do three things: (a) On the assumption that the tree-stump effect might quickly fade away, pick a level above the bottom IQ level—I will choose the IQ range of 55–70; (b) posit that the 7.38-point score difference between the WISC–R and WAIS–R at that level really does represent an IQ loss over 6 years—by 16- to 17-year-olds; and (c) calculate a progression from ages 5 to 15 in which IQ gains turn into IQ losses, such that the gain for all ages collectively remains at 2.4 points.

As Table 4 shows, we must posit a progression in which a 12.20-point gain (over 6 years) for 5-year-olds gradually turns into a 7.38-point loss by age 16. Such values are simply too huge to be credible: No existent data yield a rate of gain or loss of even half their magnitude. Table 4 also includes WISC gains by age from actual data: A reasonable gain of about 2 points (over 6 years) shows no tendency to vary with age. The contrast reinforces the point: WAIS–R data at low-IQ levels tell us something about different scoring conventions but nothing about IQ trends over time.

This leads to a piece of unfinished business: Reverting to Table 2 and its WAIS to WAIS–R data, because this comparison includes the tainted WAIS–R, there now is reason to be suspicious of the values at very low-IQ levels. A look at the WAIS manual shows that this suspicion is justified. An IQ of 48 requires a standard score of 16, which entails a raw score of 43, or at least 22 items correct (Wechsler, 1955, pp. 79, 101)—which means that one has to do better than a tree stump to score even a modest IQ. It is hardly surprising that these data show IQ losses up to the 70–80 score range.

In sum, one must keep in mind the point of the test comparisons. They help to estimate how much test norms have become obsolete (over time) at low IQ levels. There are no scoring convention problems with the low-IQ WISC data, and it vindicates a rate of 0.3 points per year. The tree-stump phenomenon plays havoc with the WAIS data. But a look at the bottom of Table 2 shows that it does so primarily at very low IQ levels. The WAIS to WAIS–R data start to make sense at an IQ of 70, and that, after all, is where adjusting IQ scores makes a difference as to whether someone is classified as mentally retarded. The most recent data, the WAIS–R to WAIS–III, are little affected by the tree-stump phenomenon: When the WAIS–III scores are adjusted by the usual 2.34 points, those data vindicate 0.3 points at all levels. An IQ of 45 on the WAIS–III requires only one right answer (Wechsler, 1997, pp. 181, 197). However, an IQ of 48 requires 19 correct, and I suspect that the fact that there is no discrepancy at the level of 55 is due to small numbers.

Where does this leave the defendants on death row? The IQ scores most relevant to their fate are those taken before the age of 18, which almost always means some version of the WISC rather than the WAIS. Because WISC scores do not have to be converted into WAIS scores, the problems that attend such conversions are irrelevant. All that has to be done is to update obsolete WISC IQs in terms of current WISC norms, and when that is done, we stand on firm ground.

As for testing postconviction defendants, sometimes we inherit WAIS–R scores from the past: Today's defendant may have been assessed on the WAIS–R when convicted of a previous offense as an adult. Those scores should be adjusted